REMARKS

Claims 7 to 14 are now pending. Claims 11 and 12 have been amended to incorporate any features of any referred to claims. No new matter has been added. Above, any amendments to the claims are shown by underlining (additions) and strikeouts (deletions). The Specification has been amended for clarification purposes. No new matter has been added. Please note that "Substitute Specification A" is the Specification of Record including any changes made by Applicant's Preliminary Amendment. A Marked Up Copy of Substitute Specification A accompanies this amendment. "Substitute Specification B" is the Specification Applicants are requesting be made of Record here. A Marked Up Copy of Substitute Specification B (compared with A) accompanies this amendment.

Applicant respectfully requests reconsideration of the present application in view of this response.

Applicant thanks the Examiner for acknowledging Applicant claim for foreign priority.

With regard to paragraph one (1) of the Office Action, the Examiner notes the various page and line numbers differences between the Specification and Applicants' Preliminary Amendment's amendments. Accordingly, to avoid any confusion or delay, attached hereto is Substitute Specification A which is the originally filed Specification incorporating the amendments of the Preliminary Amendment. A Marked Up Copy of the Substitute Specification A is attached hereto showing any additions (underlining) and deletions (strikeouts).

With regard to paragraphs two (2) and three (3) of the Office Action, Applicant thanks the Examiner for considering the IDS, including the German search report.

With regard to paragraph four (4) of the Office Action, Applicant thanks the Examiner for noting the typo in Applicant's Declaration and for waiving any action by Applicant at this time.

With regard to paragraph five (5) of the Office Action, Applicant thanks the Examiner for accepting Applicant's proposed substitute drawings.

With regard to paragraph six (6) of the Office Action, the title of the Application has been objected to for not being sufficiently descriptive of the invention. Applicant respectfully submits that the current title: Method for Determining Speech Quality Using Objective Measures is sufficiently descriptive. However, Applicant has submitted a new title above. No new matter has been added.

With regard to paragraph seven (7) of the Office Action, the Abstract was objected to for its last sentence. Applicant has amended the Abstract as a part of the Amendments above to the Specification (including Abstract). No new matter has been added. The Abstract was also objected to for its first two sentences as comparing the invention with the prior art. Applicant respectfully submits that the Abstract does not compare the invention with the prior art. Instead, the first sentence indicates an objective of the method being stated. And, the second sentence indicates at least one additional feature of the method. Accordingly, Applicant respectfully requests that the amendment to the Abstract be accepted and that the Abstract as is is approved by the Examiner. Withdrawal of the objection to the Abstract is respectfully requested.

With regard to paragraph eight (8) of the Office Action, the Specification was objected to for not including a Brief Description of the Drawings. Applicant has amended the Specification above and a Brief Description of the Drawings has been added. No new matter has been added. Accordingly, Applicant respectfully requests that the amendment to the Specification be accepted and that the Specification is approved by the Examiner. Withdrawal of the objection to the Specification is respectfully requested.

With regard to paragraphs nine (9), seventeen (17) and eighteen (18) of the Office Action, Applicant thanks the Examiner for allowing claims 11 to 14, provided they are rewritten to include all of the limitations of the corresponding base claims. Accordingly, Applicant has amended claims 11 to 14 above. Withdrawal of the objections and allowance of claims 11 to 14 is respectfully requested.

With regard to paragraph ten (10) of the Office Action, Applicant thanks the Examiner for noting Applicant's typographical error in claim 11, line 3 which should read "characteristic" not "characteristics." Applicant has amended claim 11 in accordance with the Examiner's suggestion.

With regard to paragraphs eleven (11) to sixteen (16) of the Office Action, claims 7 to 10 were rejected under 35 U.S.C. § 103(a) as being unpatentable over International Application Publication No. WO 96/28952 to Beerends et al ("Beerends reference") in view of U.S. Patent No. 5,621,854 to Hollier ("Hollier reference").

The Beerends reference purportedly describes a device for determining the quality of an output signal having a first series circuit for receiving the output signal and a second series circuit for receiving the reference signal and generating an objective quality signal by a combining circuit coupled to the two series circuits. Abstract. The Beerends reference refers to the correlation

between the quality signal and a subjective quality signal can be improved by coupling a converting arrangement to a series circuit for converting at least two signal parameters into a third signal parameter, and by coupling a discounting arrangement to the converter arrangement for discounting the third signal parameter at the combining circuit. Id.

The Hollier reference purportedly describes a method and apparatus for objective speech quality measurements of telecommunication equipment. Title. The Hollier reference refers to a telecommunications testing apparatus having a signal generator which generates a speech-like synthetic signal, which is supplied to an analyzer. Abstract. The analyzer derives a measure of the excitation of the human auditory system generated by both the undistorted test signal and the distorted test signal. Abstract. The difference between the two excitations is then calculated and a measure of the loudness of the difference is derived which is found to indicate to a high degree of accuracy the human subjective response to the distortion introduced by the telecommunications system. Abstract.

Claim 7 is directed to a method for determining speech quality and recites:

calculating a speech quality characteristic value by comparing respective spectral short-time properties of an assessed speech signal and of a reference speech signal;

prior to the comparing the respective spectral short-time properties, reducing differences in respective mean spectral envelopes of the assessed speech signal and of the reference speech signal by weighting spectral short-time properties of the assessed speech signal and the reference speech signal in a predetermined number of time segments using a spectral weighting function so as to include differences in the respective mean spectral envelopes in the speech quality characteristic value to a limited extent, the spectral weighting function being calculated from the respective mean spectral envelopes; and

calculating a respective intensity value for each of a plurality of frequency bands in a signal segment respectively for the assessed speech signal and the reference speech signal using variable limits for the frequency bands so that a respective difference between each calculated respective intensity of the assessed speech signal and the reference speech signal is reduced.

Neither the Beerends reference nor the Hollier reference, alone or in combination, recite all of the features of claim 7 in the manner shown. Neither cited reference describes or suggests that prior to the comparing the respective spectral short-time properties, one should reducing differences in respective mean spectral envelopes of the assessed speech signal and of the reference speech

signal by weighting spectral short-time properties of the assessed speech signal and the reference speech signal in a predetermined number of time segments using a spectral weighting function so as to include differences in the respective mean spectral envelopes in the speech quality characteristic value to a limited extent, the spectral weighting function being calculated from the respective mean spectral envelopes. Further, neither reference describes or suggests calculating a respective intensity value for each of a plurality of frequency bands in a signal segment respectively for the assessed speech signal and the reference speech signal using variable limits for the frequency bands so that a respective difference between each calculated respective intensity of the assessed speech signal and the reference speech signal is reduced. Instead, for example, the Beerends reference describes using a first signal parameter which is represented by means of a time spectrum and a Bark spectrum which is then converted into a compressed arrangement. The reference continues in its description, but does not appear to concern itself with the weighting spectral short-time properties of the speech signals in a predetermined number of time segments using a spectral weighting function so as to include differences in the respective mean spectral envelopes in the speech quality characteristic value (and the spectral weighting function is calculated from the mean spectral envelopes). For these and other reasons, the Beerends reference, alone or in combination with the Hollier reference, do not appear to render obvious claim 7 under 35 U.S.C. § 103(a) and withdrawal of the rejection is respectfully requested.

Since claims 8 to 10 depend, directly or indirectly from claim 7, claims 8 to 10 are allowable for at least the same reasons as claim 7.

Moreover, to reject a claim as obvious under 35 U.S.C. § 103, the prior art must disclose or suggest each claim element and it must also provide a motivation or suggestion for combining the elements in the manner contemplated by the claim. (See Northern Telecom, Inc. v. Datapoint Corp., 908 F.2d 931, 934 (Fed. Cir. 1990), cert. denied, 111 S. Ct. 296 (1990); In re Bond, 910 F.2d 831, 834 (Fed. Cir. 1990)).

The Federal Circuit in the case of In re Kotzab has made plain that even if a claim concerns a "technologically simple concept" -- which is not even the case here, there still must be some finding as to the "specific understanding or principle within the knowledge of a skilled artisan" that would motivate a person having no knowledge of the claimed subject matter to "make the combination in the manner claimed", stating that:

> In this case, the Examiner and the Board fell into the hindsight trap.

> The idea of a single sensor controlling multiple valves, as opposed to multiple sensors controlling multiple valves, is a technologically simple concept. **With this simple concept in mind, the Patent and Trademark Office found prior art statements that in the abstract appeared to suggest the claimed limitation. But, there was no finding as to the specific understanding or principle within the knowledge of a skilled artisan that would have motivated one with no knowledge of Kotzab's invention to make the combination in the manner claimed.** In light of our holding of the absence of a motivation to combine the teachings in Evans, we conclude that the Board did not make out a proper *prima facie* case of obviousness in rejecting [the] claims . . . under 35 U.S.C. Section 103(a) over Evans.

(See In re Kotzab, 55 U.S.P.Q.2d 1313, 1318 (Federal Circuit 2000) (citations omitted, italics in original, emphasis added)). Here again, there have been no such findings.

In addition, with respect to the above-identified application, Applicants request some sort of evidence and/or affidavit from the Patent Office regarding the Patent Office's assertions of what it suggests is obvious to one of ordinary skill in the art. See Office Action, page 3.

No motivation or suggestion for combining the elements in the manner contemplated by claim 7 is shown in the Beerends nor the Hollier references, alone or in combination.

Accordingly, it is respectfully submitted that the rejection of claims 7 to 10 under 35 U.S.C. § 103(a) over the Beerends reference in view of the Hollier reference should be withdrawn.

CONCLUSION

In view of all of the above, it is believed that the objections to the Specification (including Abstract), and Title have been overcome, rejections of claims 7 to 10, under 35 U.S.C. § 103(a) have been obviated, and that all currently pending claims 7 to 14 are allowable. It is therefore respectfully requested that the rejections be reconsidered and withdrawn, and that the present application issue as early as possible.

If it would further allowance of the present application, the Examiner is invited to contact the undersigned at the contact information shown below.

Respectfully submitted, By: Linda Shudy Reg No. 47084

Dated: April 8, 2004

By: Richard L. Mayer
Richard L. Mayer
(Reg. No. 22,490)

KENYON & KENYON
One Broadway
New York, New York 10004
Tel. (212) 425-7200
Fax (212) 425-5288

CUSTOMER NO. 26646

METHOD FOR DETERMINING SPEECH QUALITY USING OBJECTIVE MEASURES

[~~Preliminary Remarks~~

]Field of the Invention

The present invention relates to a method for determining speech quality using objective measures, in which characteristic values for determining speech quality are derived by comparing properties of a speech signal to be assessed to properties of a reference speech signal[~~ (~~], or undisturbed signal[~~)~~].

[~~Usually~~]Related Technology

Typically, the quality of speech signals is determined through auditory ("subjective") tests by test persons.

The aim of objective methods for determining speech quality is to ascertain, with the aid of suitable calculation methods, characteristic values from the properties of the speech signal to be assessed, the characteristic values describing the speech quality of the speech signal to be assessed, without having to resort to the judgments of test persons.

The calculated characteristic values and the underlying method for determining speech quality using objective measures are regarded as acknowledged if a high correlation with the results of auditory reference tests is achieved. Consequently, the speech-quality values obtained by auditory tests represent the target values which are to be achieved by objective methods.

[~~Related Art~~

]Known methods for determining speech quality using objective measures are based

on

a comparison of a reference speech signal to the speech signal to be assessed. In this context, the reference speech signal and the speech signal to be assessed are segmented into short time segments. The spectral properties of the two signals are compared in these segments.

Various approaches and models are used to calculate the spectral short-time properties. Generally, the signal intensity is calculated in frequency bands whose width becomes greater with increasing mid-frequency. Examples of such frequency bands are the known third-octave bands or frequency groups according to Zwicker (published in Zwicker, E.: "*Psychoakustik*" ["Psychoacoustics"], Berlin: Springer Publishing House, 1982).

The spectral intensity representation thus calculated for each time segment considered can be viewed as a series of numerical values, in which the number of individual values corresponds to the number of frequency bands used, the numerical values themselves represent the calculated intensity values, and a consecutive index of the frequency bands describes the sequence of the numerical values.

In the methods presently known for determining speech quality using objective measures, the limits of the frequency bands utilized are kept constant on the frequency axis.

In each time segment under consideration, the calculated intensities of the speech signal to be assessed and of the reference speech signal are compared to each other in each band. The difference of both values, or the similarity of the two resulting spectral intensity representations, constitutes the basis for the calculation of a quality value (See Fig. 1).

Such methods were developed in particular for the qualitative assessment of speech in

telephone applications. Examples thereof are the publications:

"*A perceptual speech-quality measure based on a psychacoustic sound representation*" (Beerends, J. G.; Stemerdink, J. A., J. Audio Eng. Soc. 42(1994)3, pp. 115-123)

"*Auditory distortion measure for speech coding*" (Wang, S; Sekey, A.; Gersho, A.: IEEE Proc. Int. Conf. acoust., speech and signal processing (1991), pp.493-496).

The presently valid ITU-T standard P.861 likewise describes such a method: "*Objective quality measurement of telephone-band speech codecs*" (ITU-T Rec. P.861, Geneva 1996).

~~[Disadvantages of Known Objective Speech-Quality Measurement Methods~~

]The use of known methods for determining speech quality using objective measures fails with respect to the reliability of the calculated quality values for certain signal properties to be assessed. Presently known methods furnish only unreliable quality values in particular when the speech signal to be assessed is impaired, such as in the case of impairments caused by speech coding methods with low bit rates or combinations of different disturbances.

In such cases, the presently known methods have the disadvantage that, given a comparison between the speech signal to be assessed and a reference speech signal, the quality characteristic value to be calculated includes differences between the two signal segments in the selected representation plane which either do not lead or scarcely lead to a qualitative impairment, not even one which is perceptible in the auditory test.

Within the framework of the transmission of speech in telephone applications that is being discussed here, frequency-band limitations and spectral deformations of the

speech signal to be assessed (caused, for example, by filter properties of the telephone device or of the transmission channel) contribute only to a limited extent to a perceived qualitative impairment.

To partially prevent such deficiencies, an attempt is made in a different approach to compensate for the linear distortions (frequency response) by a correction filter or a power-transmission function ([~~published in:~~]See "*A new approach to objective quality-measures based on attribute-matching*", Halka, U.; Heute, U., Speech communication, 11(1992)1, pp.15-30). However, the use of this method is disadvantageous in the case of nonlinear and time-invariant transmission, since the compensation function thus calculated no longer exclusively describes the spectral deformations of the signal to be assessed.

In known methods, displacements of spectral short-time maxima ("formant displacements") in the signal under test in relation to the reference speech signal caused, for example, by coding systems with low bit rates, lead to large differences in the spectral intensity representations and therefore have a great influence on the calculated quality value. However, investigations have revealed that, in an auditory speech-quality test, these displacements of spectral short-time maxima have only a limited influence on the quality judgment.[

~~Object~~]

[~~T~~]Summary of the Invention

An object of the invention is to reduce the influence of spectral limitations and deformations of the speech signal to be assessed, as well as the influence of displacements of spectral short-time maxima, prior to comparing the spectral properties of a signal to be tested to a reference speech signal, and prior to the calculation of a quality value using objective methods.

~~**Achievement**~~

MARKED UP COPY OF THE SUBSTITUTE SPECIFICATION A

In contrast to known approaches, [~~in~~]according to the present invention[~~ described here~~], a spectral weighting function is generated which is based on mean spectral envelopes, e.g., the mean spectral power density, of the speech signal to be assessed and the reference speech signal. This permits the use of the method in the case of nonlinear and time-variant transmission as well.

The spectral weighting function is calculated from the quotients of the given values of the mean spectral power density of the signal to be assessed $Phi_y(f)$ and that of the input signal of the transmission system $Phi_x(f)$, such that the weighting function can be described via

$$W_T(f) = a(f) \cdot (Phi_y(f) / Phi_x(f)).$$

The assessment function a(f) can weight the weighting function $W_T(f)$ differently over the range of effect, being constant at 1 in the simplest case.

The spectral weighting function $W_T(f)$ thus calculated brings the mean spectral envelopes of the speech signal to be assessed and the reference speech signal closer to each other, so that differences of the two spectral envelopes are included only to a reduced extent in the calculated quality value.

The spectral weighting function $W_T(f)$ can be applied, firstly, to the reference speech signal. In this context, the reference speech signal, in its mean spectral power density, is made to approximate the signal to be assessed (Fig. 2a).

Secondly, the spectral weighting function can be applied, inverted, to the signal to be assessed. The distortion of the latter is thereby eliminated and, with regard to its mean spectral power density, it is made to approximate the reference speech signal (Fig. 2b).

A further [~~part~~]aspect of the present invention relates to the correction of

displacements of spectral short-time maxima which are caused by the transmission systems.

The intensity is integrated for each time segment in frequency bands. The result is a series of intensity values for each spectral representation of a signal segment, each individual value representing the intensity in a frequency band. In this connection, the displacements of spectral short-time maxima may lead to different calculated intensities in the frequency bands of the reference speech signal and the speech signal to be assessed.

These differences in the spectral intensity representations - caused by displacements of spectral short-time maxima - can be reduced by a variable arrangement of the frequency bands on the frequency axis. In contrast to the constant band limits in known methods, the band limits are displaced on the frequency axis. However, the number of frequency bands and their index remain constant. In an optimization loop, those band limits are then accepted at which the two resulting spectral representations of speech signal to be assessed and reference speech signal exhibit maximum similarity, or whose difference is minimal. This optimization is carried out for all bands in all time segments under consideration.

The use of variable band limits to calculate the spectral intensity representation is not restricted only to the signal in which the described spectral weighting function $W_T(f)$ is also used, but may also be applied to the other respective signal and even to both signals (see Fig. 2a and 2b).

~~[Exemplary Embodiment:~~

~~A special exemplary embodiment is shown by an implementation according]~~Brief Description of the Drawings

Fig. 1 shows a flow chart depicting a prior art calculation of a quality value;

Fig. 2a shows a flow chart depicting a calculation of a quality value using a spectral

weighting function;

Fig. 2b shows a flow chart depicting a calculation of a quality value using an inverted spectral weighting function; and

Fig. 3 shows a flow chart depicting a calculation of a Telecommunication Objective Speech Quality Assessment (TOSQA) using a spectral weighting function.

Detailed Description

An embodiment of the present invention is now described with reference to Fig. 3, which ~~[is known as]~~shows a flowchart depicting a calculation of a so-called TOSQA (Telecommunication Objective Speech Quality Assessment). In this case, an expanded preprocessing of the reference speech signal is carried out.

~~[In specification of]~~Following the general implementations according to Fig. 2a and 2b, ~~[s]~~but with more specificity, reference speech signal 2 and the speech signal to be assessed 4 are segmented (see blocks 6 and 8, respectively). Speech pauses are detected here by a speech-pause detector (see block 10) and are not included in the quality measure. Likewise, ~~[the]~~reference speech signal 2 and~~[the]~~ speech signal to be assessed 4 are filtered with a 300 ... 3400 Hz bandpass filter (see blocks 14 and 16, respectively), and there is also filtering to the frequency response of a telephone handset (see blocks 18 and 20, respectively). The weighting function $W_T(f)$ is applied to the reference speech signal before the bandpass filtering (see block 12). The integration of the spectral power density is carried out in frequency groups which represent the basis for the calculation of the specific loudness (see blocks 22 and 24, respectively).

However, the integration in frequency groups is *not* carried out in fixed frequency-group limits, but with the variable frequency-group limits described in the present invention. The calculated signal powers in the frequency groups thus modified form the basis for the intensity calculation. Use was made here of a model for calculating the specific loudness according to Zwicker, an aurally compensated intensity representation (published in Zwicker, E.: "*Psychoakustik*" ["Psychoacoustics"],

Berlin: Springer Publishing House, 1982), which is hereby incorporated by reference herein.

As an addition to the general approach, the calculated loudness patterns are supplemented by an error assessment function (see block 26). The calculated quality value TOSQA is formed via a mean value of the correlation coefficients of the specific loudness for each short time segment under consideration over the number of evaluated speech segments (see block 28).

~~Patent Claims~~ WHAT IS CLAIMED IS:

Abstract

~~Known methods for instrumental voice quality evaluation based on comparing signal intensities of the voice signal to be evaluated with a reference voice signal do not optimally evaluate spectral distortions in the voice signal to be evaluated so that quality evaluation is unreliable. Moreover, by integrating the signal intensity in the frequency bands with constant band limits, certain falsifications of the voice signal to be evaluated, such as those caused, for instance, by coding systems with lower bit rates, are erroneously evaluated. In]~~In a method for determining speech quality using an objective measure, in order to enhance prediction reliability of the evaluated quality parameters, distortions of the mean spectral envelope are extensively corrected with a weighting function $W_T(f)$ before comparing spectral properties. ~~[On the other hand]~~Additionally, the fixed band limits for integration of spectral power density are suppressed and other band limits are searched for instead in a predetermined optimization area in which the resulting spectral intensity representations of the voice signal to be evaluated and the reference voice signal have maximum similarity. The solutions described can supplement known methods and can be incorporated into their structures.

[2345/127]

METHOD FOR DETERMINING SPEECH QUALITY USING OBJECTIVE MEASURES

Field of the Invention

The present invention relates to a method for determining speech quality using objective measures, in which characteristic values for determining speech quality are derived by comparing properties of a speech signal to be assessed to properties of a reference speech signal, or undisturbed signal.

Related Technology

[~~Typically, the~~]The quality of speech signals [~~is~~]may be determined through auditory ("subjective") tests by test persons.

[~~The aim of o~~]Objective methods for determining speech quality[~~ is to~~] ascertain, with the aid of suitable calculation methods, characteristic values from the properties of the speech signal to be assessed, the characteristic values describing the speech quality of the speech signal to be assessed, without having to resort to the judgments of test persons.

The calculated characteristic values and the underlying method for determining speech quality using objective measures are regarded as acknowledged if a high correlation with the results of auditory reference tests is achieved. Consequently, the speech-quality values obtained by auditory tests represent the target values which are to be achieved by objective methods.

[~~Known~~]Available methods for determining speech quality using objective measures are based on[

]_a comparison of a reference speech signal to the speech signal to be assessed. In this context, the reference speech signal and the speech signal to be assessed are segmented into short time segments. The spectral properties of the two signals are compared in these segments.

MARKED UP COPY OF THE SUBSTITUTE SPECIFICATION B

Various approaches and models are used to calculate the spectral short-time properties. Generally, the signal intensity is calculated in frequency bands whose width becomes greater with increasing mid-frequency. Examples of such frequency bands are the known third-octave bands or frequency groups according to [~~Zwicker (published in Zwicker, E.: "*Psychoakustik*"~~]reference "Psychoakustik" [[~~"Psychoacoustics"~~]"Psychoacoustics"], by E. Zwicker, Berlin: Springer Publishing House, 1982[~~)~~].

The spectral intensity representation thus calculated for each time segment considered can be viewed as a series of numerical values, in which the number of individual values corresponds to the number of frequency bands used, the numerical values themselves represent the calculated intensity values, and a consecutive index of the frequency bands describes the sequence of the numerical values.

In [~~the~~]available methods[~~ presently known~~] for determining speech quality using objective measures, the limits of the frequency bands utilized are kept constant on the frequency axis.

In each time segment under consideration, the calculated intensities of the speech signal to be assessed and of the []reference speech signal are compared to each other in each band. The difference of both values, or the similarity of the two resulting spectral intensity representations, constitutes the basis for the calculation of a quality value ([~~S~~]see Fig. 1).

Such methods were developed [~~in particular ~~]for the qualitative assessment of speech in telephone applications. [~~E~~]Some examples [~~thereof ~~]are illustrated in the [~~publications~~]following references:[

]-"*A perceptual speech-quality measure based on a psychacoustic sound representation*," [~~(~~]by J.G. Beerends[~~,~~] and J.[~~G~~]A.[~~,~~] Stemerdink, J.[~~A., J.~~] Audio Eng. Soc. 42(1994)3, pp. 115-123[~~)~~]

]:"*Auditory distortion measure for speech coding*," [~~(~~]by S. Wang, [~~S,~~]A. Sekey, and A.[~~,~~] Gersho,[~~A.:~~] IEEE Proc. Int. Conf. acoust., speech and signal processing (1991), pp.493-496[~~).~~]

~~The presently valid~~]; and ITU-T standard P.861[~~ likewise describes such a method:~~], *"Objective quality measurement of telephone-band speech codecs,"* [~~(~~]ITU-T Rec. P.861, Geneva 1996[~~)~~].

The use of [~~known~~]available methods for determining speech quality using objective measures fails with respect to the reliability of the calculated quality values for certain signal properties to be assessed. Presently [~~known~~]available methods furnish only unreliable quality values in particular when the speech signal to be assessed is impaired, such as in the case of impairments caused by speech coding methods with low bit rates or combinations of different disturbances.

In such cases, the presently [~~known~~]available methods have the disadvantage that, given a comparison between the speech signal to be assessed and a reference speech signal, the quality characteristic value to be calculated includes differences between the two signal segments in the selected representation plane which either do not lead or scarcely lead to a qualitative impairment, not even one which is perceptible in the auditory test.

Within the framework of the transmission of speech in telephone applications that is being discussed here, frequency-band limitations and spectral deformations of the speech signal to be assessed (caused, for example, by filter properties of the telephone device or of the transmission channel) contribute only to a limited extent to a perceived qualitative impairment.

To partially prevent such deficiencies, an attempt is made in a different approach to compensate for the linear distortions (frequency response) by a correction filter or a power-transmission function[~~ (~~]. See, e.g., "*A new approach to objective quality-measures based on attribute-matching*", by U. Halka[~~,~~] and U.[~~;~~] Heute[~~, U.~~], Speech communication, 11(1992)1, pp.15-30[~~)~~]. However, the use of this method is disadvantageous in the case of nonlinear and time-invariant transmission, since the compensation function thus calculated no longer exclusively describes the spectral deformations of the signal to be assessed.

In [~~known~~]available methods, displacements of spectral short-time maxima ("formant

displacements") in the signal under test in relation to the reference speech signal caused, for example, by coding systems with low bit rates, lead to large differences in the spectral intensity representations and therefore have a great influence on the calculated quality value. However, investigations have revealed that, in an auditory speech-quality test, these displacements of spectral short-time maxima have only a limited influence on the quality judgment.

Summary of the Invention

An object of the invention is to reduce the influence of spectral limitations and deformations of the speech signal to be assessed, as well as the influence of displacements of spectral short-time maxima, prior to comparing the spectral properties of a signal to be tested to a reference speech signal, and prior to the calculation of a quality value using objective methods.

In contrast to ~~[known]~~available approaches, according to the present invention, a spectral weighting function is generated which is based on mean spectral envelopes, e.g., the mean spectral power density, of the speech signal to be assessed and the reference speech signal. This permits the use of the method in the case of nonlinear and time-variant transmission as well.

The spectral weighting function is calculated from the quotients of the given values of the mean spectral power density of the signal to be assessed $Phi_y(f)$ and that of the input signal of the transmission system $Phi_x(f)$, such that the weighting function can be described via

$$W_T(f) = a(f) \cdot (Phi_y(f) / Phi_x(f)).$$

The assessment function a(f) can weight the weighting function $W_T(f)$ differently over the range of effect, being constant at 1 in the simplest case.

The spectral weighting function $W_T(f)$ thus calculated brings the mean spectral envelopes of the speech signal to be assessed and the reference speech signal closer to each other, so that differences of the two spectral envelopes are included only to a reduced extent in the calculated quality value.

The spectral weighting function $W_T(f)$ can be applied, firstly, to the reference speech signal. In this context, the reference speech signal, in its mean spectral power density, is made to approximate the signal to be assessed (Fig. 2a).

Secondly, the spectral weighting function can be applied, inverted, to the signal to be assessed. The distortion of the latter is thereby eliminated and, with regard to its mean spectral power density, it is made to approximate the reference speech signal (Fig. 2b).

A further aspect of the present invention relates to the correction of displacements of spectral short-time maxima which are caused by the transmission systems.

The intensity is integrated for each time segment in frequency bands. The result is a series of intensity values for each spectral representation of a signal segment, each individual value representing the intensity in a frequency band. In this connection, the displacements of spectral short-time maxima may lead to different calculated intensities in the frequency bands of the reference speech signal and the speech signal to be assessed.

These differences in the spectral intensity representations - caused by displacements of spectral short-time maxima - can be reduced by a variable arrangement of the frequency bands on the frequency axis. In contrast to the constant band limits in known methods, the - band limits are displaced on the frequency axis. However, the number of frequency bands and their index remain constant. In an optimization loop, those band limits are then accepted at which the two resulting spectral representations of speech signal to be assessed and reference speech signal exhibit maximum similarity, or whose difference is minimal. This optimization is carried out for all bands in all time segments under consideration.

The use of variable band limits to calculate the spectral intensity representation is not restricted only to the signal in which the described spectral weighting function $W_T(f)$ is also used, but may also be applied to the other respective signal and even to both signals (see Fig. 2a and 2b).

Brief Description of the Drawings

Fig. 1 shows a flow chart depicting a prior art calculation of a quality value[~~;~~].

Fig. 2a shows a flow chart depicting a calculation of a quality value using a spectral weighting function[~~;~~].

Fig. 2b shows a flow chart depicting a calculation of a quality value using an inverted spectral weighting function[~~; and~~].

Fig. 3 shows a flow chart depicting a calculation of a Telecommunication Objective Speech Quality Assessment (TOSQA) using a spectral weighting function.

Detailed Description

[~~A~~]Fig. 3 shows an embodiment [~~of~~]according to the present invention[~~ is now described with reference to Fig. 3~~],[~~ which shows~~] showing a flowchart depicting a calculation of a so-called TOSQA (Telecommunication Objective Speech Quality Assessment). In this case, an expanded preprocessing of the reference speech signal is carried out.

Following the general implementations according to Fig. 2a and 2b, but with more specificity, reference speech signal 2 and the speech signal to be assessed 4 are segmented (see blocks 6 and 8, respectively). Speech pauses are detected here by a speech-pause detector (see block 10) and are not included in the quality measure. Likewise, reference speech signal 2 and speech signal to be assessed 4 are filtered with a 300 ... 3400 Hz bandpass filter (see blocks 14 and 16, respectively), and there is also filtering to the frequency response of a telephone handset (see blocks 18 and 20, respectively). The weighting function $W_T(f)$ is applied to the reference speech signal before the bandpass filtering (see block 12). The integration of the spectral power density is carried out in frequency groups which represent the basis for the calculation of the specific loudness (see blocks 22 and 24, respectively).

However, the integration in frequency groups is []not carried out in fixed frequency-group limits, but with the variable frequency-group limits described in the present invention. The calculated signal powers in the frequency groups thus modified form the basis for the intensity calculation. Use was made here of a model for calculating the specific loudness according to Zwicker, an aurally compensated intensity representation ([~~published in Zwicker, E.:~~]see "*Psychoakustik*" ["Psychoacoustics"], by E. Zwicker, Berlin: Springer Publishing House, 1982), which is hereby incorporated by reference herein.

As an addition to the general approach, the calculated loudness patterns are supplemented by an error assessment function (see block 26). The calculated quality value TOSQA is formed via a mean value of the correlation coefficients of the specific loudness for each short time segment under consideration over the number of evaluated speech segments (see block 28).

WHAT IS CLAIMED IS:

Abstract

In a method for determining speech quality using an objective measure, in order to enhance prediction reliability of the evaluated quality parameters, distortions of the mean spectral envelope are extensively corrected with a weighting function $W_T(f)$ before comparing spectral properties. Additionally, the fixed band limits for integration of spectral power density are suppressed and other band limits are searched for instead in a predetermined optimization area in which the resulting spectral intensity representations of the voice signal to be evaluated and the reference voice signal have maximum similarity. [~~The solutions described can supplement known methods and can be incorporated into their structures.~~]

MARKED UP COPY OF THE SUBSTITUTE SPECIFICATION B

NY01 657303 v 1